

Mrežni izvori Sveučilišta u Göteborgu i Instituta za njemački jezik

Daria Lazić i Josip Mihaljević



Jezikoslovnne
RASPRAVE

Sveučilište u Göteborgu

11. do 15. veljače 2019. (Daria Lazić)



Sveučilište u Göteborgu

- Odsjek za švedski jezik:
 - Språkbanken – centar za jezične tehnologije
 - Centar za leksikologiju i leksikografiju
 - uredništvo rječnika Švedske akademije



GÖTEBORGS
UNIVERSITET

Språkbanken

- osnovan 1975.
- cilj: prikupljanje i razvoj jezičnih izvora te njihovo stavljanje na raspolaganje znanstvenoj i široj javnosti
 - korpusi, leksikoni, jezični alati
- od 2018. – dio Nacionalne jezične banke čiji je cilj razvoj nacionalne e-infrastrukture za istraživanja temeljena na jezičnim podacima
 - tri odjela: SBSAM, SBTAL, **SBTEXT**
 - Sveučilište u Göteborgu: SBTEXT – obrada pisanoga jezika
 - projekt financira švedska zaklada za znanost i uključene visokoobrazovne i znanstvene ustanove

Språkbanken

- 30-ak zaposlenih: stručnjaci i programeri
- korisnici: uglavnom lingvisti i istraživači u području jezičnih tehnologija, ali i humanističkih i društvenih znanosti općenito
 - kontakt s korisnicima: savjetovanje i podrška
- projekti obuhvaćaju razvoj jezičnih tehnologija za uporabu u:
 - istraživanju jezičnih varijacija (analiza povijesnih podataka, računalna morfologija, poredbena lingvistika)
 - analizi sadržaja (razlučivanje i promjena značenja, analiza sentimenta, rano otkrivanje demencije)
 - učenju jezika (analiza složenosti tekstova, računalno potpomognuto učenje)
- rezultati istraživanja dostupni su na mrežnim platformama s otvorenim pristupom

Korp



Korp

[User manual](#)

Search in the corpus collections

corpora count	462
token (total)	15 222 714 020
sentences (total)	1 060 995 330

- program za pretraživanje korpusa
- sadržava velik broj korpusa:
 - suvremeni
 - povijesni
 - paralelni
- mogućnost prilagodbe za druge jezike

Modern | Parallel | Old Swedish | Litteraturbanken | Kubhist | Old texts | More ▾



125 of 237 corpora selected — 2.08G of 13.26G tokens

Simple Extended

in order and also as init

KWIC: hits per page: 25 ▾

1850 1900 1950 2000

Select all
 Select none

- ▾ Akademiska texter (2)
- ▾ August Strindberg (2)
- ▾ Finlandssvenska texter (56)
- ▾ Medicinska texter (13)
- ▾ Myndighetstexter (24)
- ▾ Skyddade korpusar (13)
- ▾ Skönlitteratur (6)
- ▾ Sociala medier (67)
- ▾ Tidningstexter (37)
- ▾ Tidskrifter (1)
- COCTALL
- Dramawebben
- Etnologiska frågelistor
- IVIP demo
- LäSBarT – Lättläst svenska och barnbokstext
- PAROLE
- Poeter.se
- Psalmboken (1937)
- SALT svenska-nederländska
- SNP 78–79 (Riksdagens snabbprotokoll)



- jednostavno, prošireno i napredno pretraživanje
- pretraživanje prema značenjima
- prikaz rezultata: KWIC, statistika, "slika riječi" (*ordbild*)

hund (substantiv)

Preposition	Attribut	hund	Efterställt Attribut	Hund	verb	Verb	hund	
1. med 5016	1. liten 2207	1. heta 248	1. bita 227	1. skaffa 1369	2. om 1054	2. söt 584	2. vara 472	2. ha 4781
3. hos 242	3. röd 589	3. skälla 88	3. dö 192	3. rasta 372	4. till 1166	4. blöt 269	4. ha 234	4. rasta ² 372
5. utan 190	5. lös ² 218	5. av ras 57	5. springa 152	5. köpa ² 596	6. efter 377	6. lös 218	6. bli 118	6. köpa 596
7. för 1194	7. gullig 176	7. springa 61	7. skaffa 94	7. skaffa sig 162	8. vad för 78	8. egen 183	8. bita 45	8. avliva 102
9. mot 223	9. fin 413	9. i bil 61	9. bli 506	9. avliva 63	10. bredvid 25	10. stackars 146	10. vilja 84	10. få 374
11. i form 20	11. galen 159	11. kunna 122	11. heta 103	11. klappa ² 114	12. åt 145	12. herrelös 74	12. i knä 47	11. klappa 114
13. till och med 21	13. död 154	13. skola ² 117	12. rädda 75	12. slita 128	14. till och med 21	14. snäll 144	14. få 95	13. lämna 244
15. ihop med 19	15. ful 133	15. i koppel 28	13. äta 87	13. döda 139	14. hitta 102	14. snäll 144	14. få 95	14. döda 139
			15. bajs 29	15. se 451	15. ihop med 19	15. ful 133	15. i koppel 28	15. se 451

Karp



Svenska | English

Choose a resource below or use the standard Karp selection.

Språk
BANKEN

SALDO

Lexical resource for Swedish language technology.

Dalin

Dalin's dictionary – 19th century Swedish

Old Swedish

Schlyter and Söderwall

Constructicon

A database of Swedish constructions.

SweFN

Swedish Framenet

Svenskt kvinnobiografiskt lexikon

A Russian constructicon

Bliss

A symbol lexicon.

- otvorena leksička infrastruktura
- obuhvaća više od 40 leksičkih izvora, od leksikona izrađenih u Språkbanken do povijesnih digitaliziranih rječnika
- *Saldo*:
 - švedski asocijativni leksikon
 - sadržava:
 - semantički leksikon
 - morfološki leksikon

Saldo

Find entries where **grundform** equals [+ or...](#)

[+ and...](#) [+ except...](#)

Hits **16**

Page: 1 / 1

SALDO ▾ 8 HITS (DISPLAYING 8)

	SENSE	LEMGRAM	PART OF SPEECH	PRIMARY	SECONDARY	CHILDREN (PRIMARY)	CHILDREN (SECONDARY)
⚙	hund	hund (noun) 🐦 2044231	noun	djur	sällskap	Karo Pompe bandhund björnhund blandrashund more...	Idefix andeögon apportera bussa ³ folkilsken more...
⚙	hund ²	hund (noun) 🐦 2044231	noun	usling			



- mrežna platforma za učenje švedskoga
- trenutačno sadržava:
 - vježbe za lingviste (npr.. određivanje vrsta riječi)
 - vježbe za učenje jezika (npr. pogađanje riječi)
 - alat za procjenu složenosti teksta za švedski kao strani jezik
 - alat za pronalaženje primjera
 - alat za anotiranje teksta

Vježbe: slušno razumijevanje

LiWriX - Listen and Write Exercise

Mode: 0 Level: A1

[Change settings \(new game\)](#)

izbor jezične razine

Correct: 1/1



zvukovni zapis

Score: 0

odgovor

provjera

barnprogram|

Check

Hide sentences

Där kan du titta på massor av gamla tv-serier , sportprogram , **b_____** och mycket annat .

De tycker att det är fel att skämta om sådana saker i ett **b_____** .

Partiet Hamas sänder **b_____** i tv-kanalen al-Aqsa .

I Sverige är det förbud med reklam för alkohol och reklam i **b_____** .

Förslaget betyder att det kan bli reklam i **b_____** och alkoholreklam i kanaler som sänds i Sverige .

Ett **b_____** för tv som vill inspirera barn till eget berättande av både enkla som allvarliga ämnen .

I Palestina visas **b_____** där självmordsbombare hyllas och det israeliska folket beskrivs som djävulen själv .

En afrikansk gatufest med Edo Bumba Band , dansgrupper , försäljning , **b_____** och fortsatt diktläsning .

Enligt reklamen fanns figuren med i ett **b_____** på en tv-kanal .

Många föräldrar i USA protesterar mot att Apple tillåter försäljning inne i **b_____** .



plugga



pomoć:

- rečenice
- početno slovo

Vježba: pogađanje riječi

Word guess

Tries: 0/7

Definition:

lämnar ifrån sig mot betalning

Hjälp:

prodati/prodavati

S Ä L J 

A B C D E F G H I K M N O P Q R T U V W X Y Z Ö Å É

What do you want to assess? ?

Learner essay

Text readability

Show all words of the following CEFR level(s) ?

- A1
- A2
- B1
- B2
- C1

Additional options ?

- Mark all potentially incorrect words
- Use Spellchecker

Edit text

Reset



Language Acquisition Reusing **Korp**

Imorgon ska jag på bio med min bästa vän .



Language Acquisition Reusing **Korp**

Imorgon ska jag på bio med min bästa vän .

Lexikaliska institutet

- istraživanja na području leksikologije i leksikografije
- leksikografija:
 - izrada rječnika i baza podataka
 - uredništvo rječnika Švedske akademije
 - švedsko uredništvo višejezičnoga rječnika *Íslex*
 - metaleksikografija

Rječnici Švedske akademije

SVENSKA AKADEMIENS
ORDBÖCKER

HEM ORDBÖCKERNA GRAMMATIK AKADEMIEN KONTAKT

Sök i tre ordböcker på en gång

Sökhjälp

Sök ord ...

SAOL



Koncentrerad ordlista med 126 000 ord

Läs mer →

SO



Ingående beskrivningar av 65 000 ord

Läs mer →

SAOB



Historisk ordbok med 500 000 ord

Läs mer →

Rječnici Švedske akademije

- *Svensk ordbok* (2009.)
- 2009. izdan u tiskanome obliku, od 2015. dostupan i kao mrežni rječnik
- 65000 natuknica
- deskriptivan pristup

- *Svenska akademiens ordlista* (2015.)
- popis riječi s osnovnim podacima
- preskriptivan pristup

SAOL i SO

SAOL

tryckår: 2015

¹hands|free [hän´dsfri långt i] oböjligt substantiv

- *en handsfree*

– Använd hellre *handfri*, se ¹*handfri*.

[Visa mer](#) 

²hands|free

[hän´dsfri långt i] oböjligt adjektiv

– Använd hellre *handfri*, se ²*handfri*.

[Visa mer](#) 

SO

tryckår: 2009

handsfree  [hän´dsfri långt i]

substantiv, ingen böjning, n-genus

ORDLED: hands·free

- anordning i anslutning till mobiltelefon som gör att mobiltelefonen inte behöver manövreras med händerna utformad t.ex. med tanke på bilförare

SYN. *handfri*

[Visa mer](#) 

Svensk ordbok: rad na novome izdanju

- naglasak na posuvremenjavanju natuknica iako se u manjoj mjeri mijenja i njihov inventar
- rječnik treba odražavati društvene promjene
 - korekcija iz rodne, nacionalne perspektive i sl.
- promjene u strukturi natuknica – razvoj opisa rekcije
- provjera semantičkoga i morfološkoga opisa – na temelju korpusa izrađenih u Språkbanken i vlastite morfološke baze na kojoj kontinuirano rade



- mogućnost pretraživanja slobodnoga teksta u aplikaciji

”opasno putovanje u unutrašnjost Konga”

SO

tryckår: 2009

tigg`are  substantiv ~n äv. vard.

tiggarn, plur. ~, best. plur. *tiggarna*

ORDLED: tigg·ar·en

- person som (huvudsakligen) livnär sig på att tigga <mest vid beskrivning av äldre el. icke-svenska förh.>

Dölj 

EXEMPEL: *tiggarfölje; tiggarstav; de blinda tiggarna vid vägkanten*

HIST.: sedan 1526

prosjak

- osoba koja se (uglavnom) uzdržava prosjačenjem <većinom u opisima starijih ili stranih okolnosti>

SO

tryckår: 2009

me`sig 🗣️ adjektiv ~t

ORDLED: mes·ig

- rädd och ynklig

Dölj ☒

EXEMPEL: *en mesig liten kille*

- äv. om handling o.d.

EXEMPEL: *mesig tjejfotboll*

- äv. bildligt, särsk. om maträtt som smakar svagt

EXEMPEL: *en mesig soppa*

HIST.: sedan 1763

plašljiv:

PRIMJER: plašljiv ženski nogomet

- islandsko-skandinavski mrežni rječnik
- jednojezični rječnik islandskoga kao temelj i prijevod na danski, švedski, dva norveška standarda, ferski i finski

ISLEX-verkefnið Leiðbeiningar Hafa samband About

ISLEX orðabókin
Stofnun Árna Magnússonar í Íslenskum fræðum

Íslenska Dansk Norsk bokmål Nynorsk Svenska Føroyskt Suomi







Orð vikunnar Myndir

Velkomin á heimasíðu veforðabókarinnar ISLEX


á ò é í ó ú ý þ æ ö Flóknari leit
à â ø

Leita

- mögúgnost odabira jednoga ili svih jezika za prikaz

veldu orðabók:      


hundur no kk

 **framburður**

→ [BEYGING](#)


1

(dýr)




→ www.fauna.is


(ættkvíslin:)


 **hunddjur**;


(einstaklingur:)

 **hund**


2




 **odjur**


 **krák**

fara í hundana


 **gå át skogen**

 **gå át helvete**

það er hundur í <honum>

 **<han> är sur**

allt fer í hund og kött

 **allting spárar ur**

hundur no kk

 **framburður**

→ [BEYGING](#)

1

(dýr)



→ www.fauna.is

 **hund**

 **hund**

 **hund**

(ættkvíslin:)

 **hunddjur**;

(einstaklingur:)

 **hund**

 **hundur**

 **koira**

- napredne mogućnosti pretraživanja: natuknice, ekvivalenti na ciljnim jezicima, slobodno pretraživanje teksta, riječi koje se slično pišu, odabir jezika...

Leitarniðurstöður fyrir 'hund'

Orðmyndir

- hund = hundur no kk

Jafnheiti

- hund 🇫🇮 : hundur no kk

Dæmi og orðasambönd

- "vera með <hund> í bandi" : band no hk
- "algengustu gæludýrin eru hundar og kettir" : algengur lo
- "nágranninn amaðist við hundinum í garðinum" : amast so
- "eiganda hunds er skylt að auðkenna hund sinn" : auðkenna so

[á](#) [ò](#) [é](#) [í](#) [ó](#) [ú](#) [ý](#) [þ](#) [æ](#) [ö](#)
[Einföld leit](#)

[ä](#) [å](#) [ø](#)

Leita

uppflettiorð
 danska

erl. jafnheiti
 norska (bókmál)

texti
 nýnorska

loðin leit
 sænska

færøyska
 finnska

- zvučkovni zapisi, slikovni materijal, povezanost s morfološkom bazom

hundur no kk

 **framburður**

 **BEYGING**

1

(dýr)



→ www.fauna.is

(ættkvíslin:)

 hunddjur;

(einstaklingur:)

 hund

Beygingarlýsing íslensks nútímamáls

Stofnun Árna Magnússonar í íslenskum fræðum

Ritstjóri Kristín Bjarnadóttir

HEIM

UM BÍN

UM BEYGINGARÐÆMIN

ORÐAFORÐINN

MÁLTÆKNI OG GÖGN

SPURNINGAR OG SVÖR

SKRAMBI

VEFTRÉ

ENGLISH

Leita að beygingarmynd

hundur

Karlkynsnafnorð

Eintala		Fleirtala			
	án greinis	með greini	án greinis	með greini	
Nf.	hundur	hundurinn	Nf.	hundar	hundarnir
Þf.	hund	hundinn	Þf.	hunda	hundana
Þgf.	hundi	hundinum	Þgf.	hundum	hundunum
Ef.	hunds	hundsins	Ef.	hunda	hundanna

Institut za njemački jezik

5. do 8. veljače 2019. (Josip Mihaljević)



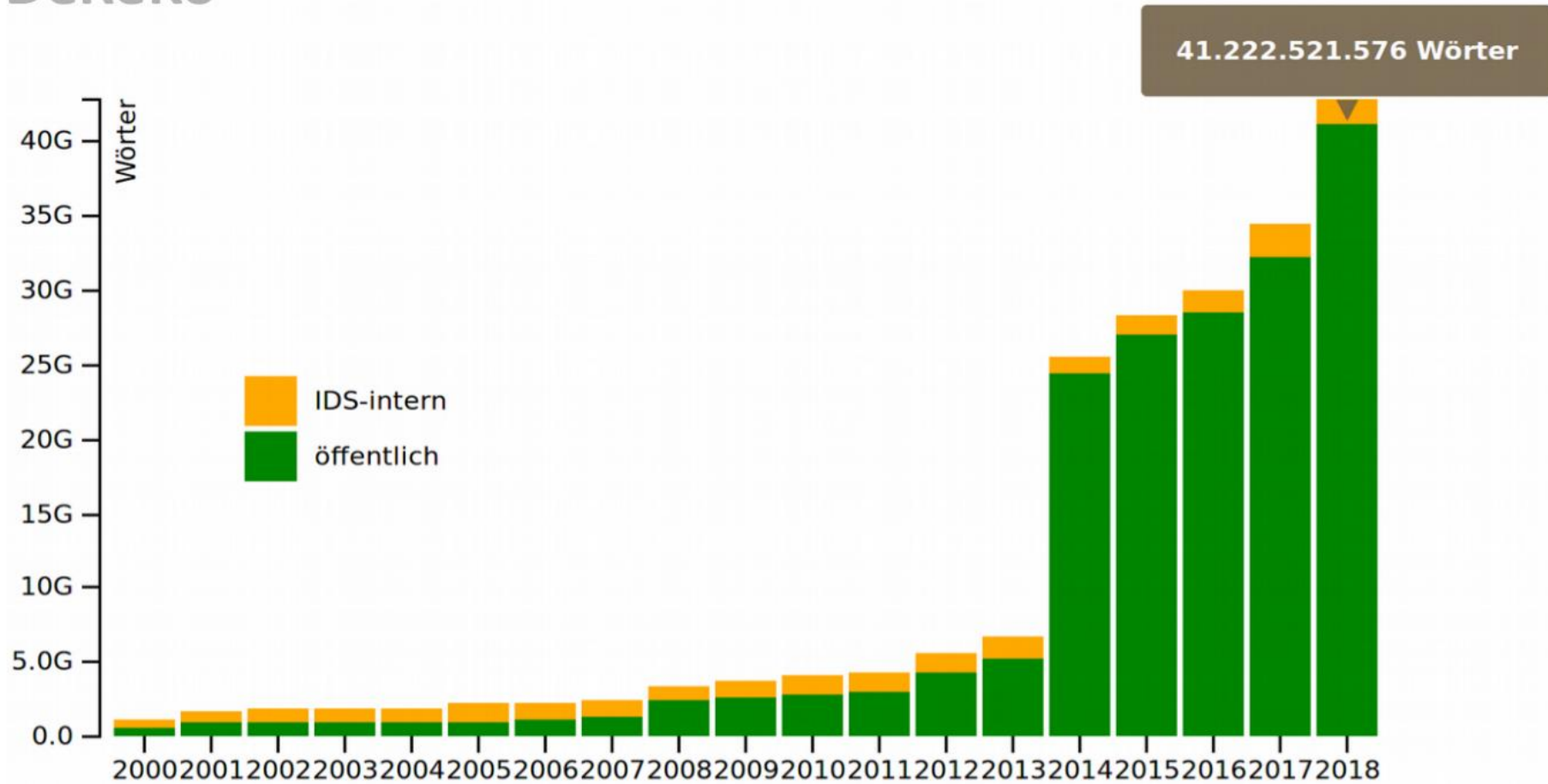
INSTITUT FÜR
DEUTSCHE SPRACHE

Osnovno o Institutu

- osnovan 1964.
- Institut nije povezan sa sveučilištem i središnja je ustanova za proučavanje suvremene i povijesne uporabe njemačkoga jezika
- uglavnom se bave dugoročnim projektima
- surađuju i s istraživačima iz drugih ustanova i sa sveučilišta
- mnogi projekti temelje se na računalnim korpusima
 - prvi računalni korpus u Institutu 1969.
 - 1983. prvi program za analizu korpusa REFER
 - 1999. razvoj i održavanje korpusa pisanoga suvremenog jezika DeReKo
 - 2004. nastaje program za pregled i analizu korpusa COSMOS II kojim se trenutačno služe

Corpora at the IDS

DeReKo



Knjižnica

- nalazi se u prizemlju zgrade (dva kata s policama)
- imaju [mrežni katalog](#) koji omogućuje pretraživanje na njemačkome i engleskome jeziku
- postoji više polja za pretraživanje te je moguće koristiti se regularnim izrazima

Multi field search

OK or Discard

Words from title Scan list

Exact title Scan list

Author Scan list

Corporate body Scan list

Subject heading Scan list

Signature Scan list







Publisher/Publishing place Scan list

Year of publication Scan list

Kind of media Index

Exact sequence of the words? No Yes

OK or Discard

#	Author ▲ ▼	Title ▲ ▼ 
<u>1</u>	<input type="checkbox"/>  Meliss, Meike	Die Entwicklung einer lexikografischen Ressource im Rahmen des Projekts LeGeDe
<u>2</u>	<input type="checkbox"/>  Meliss, Meike	Die Entwicklung einer lexikografischen Ressource im Rahmen des Projekts LeGeDe
<u>3</u>	<input type="checkbox"/>  Möhrs, Christine	Syntagmatische Verwendungsmuster in einsprachigen deutschen Wörterbüchern (Arbeiten und Materialien zur deutschen Sprache ; Band 50)
<u>4</u>	<input type="checkbox"/>  Storjohann, Petra	Das "elexiko"-Wörterbuch im Wandel - typische Verwendungen neu beleuchtet
<u>5</u>	<input type="checkbox"/>  Möhrs, Christine	Syntagmatische Verwendungsmuster in einsprachigen deutschen Wörterbüchern (Arbeiten und Materialien zur deutschen Sprache ; Band 50)

Knjižnica

- knjižnica sadržava i vanjske poveznice na druge mrežne rječnike, knjižnice, enciklopedije, jezične portale i baze podataka

The screenshot shows the Langenscheidt Online-Wörterbuch interface. The search bar contains 'pas' and the search button is labeled 'Suchen'. The results are categorized by source:

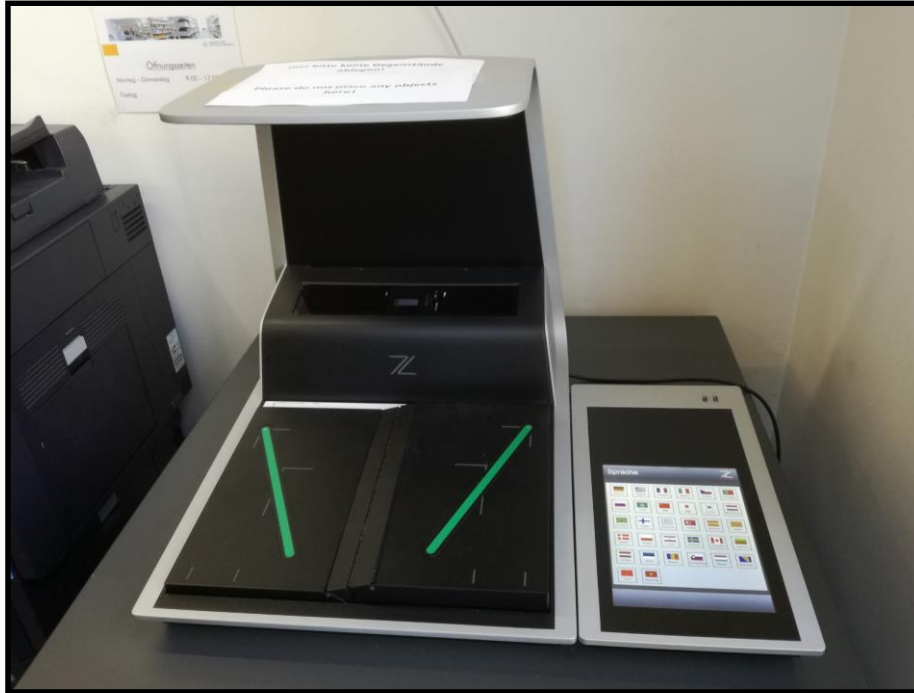
- 78 Treffer** (78 hits) from **Technik Fachwörterbuch**:
 - pas** < mwhcl > (driver steering effort is reduced; usu. by hydraulics), power assisted steering, power steering, boosted steering, p/steering • **Servolenkung f**; **Hilfskraftlenkung f**; **Lenkhilfe f**
- Größelwörterbuch**:
 - pas** Fr. - s
 - 1 right to go first **Vortritt m**
 - 2 Ballett **Pas m**, **Tanzschritt m**
- Dictionary**:
 - pas**¹ (psa; gen pl pāsā) **Hund m**
 - morski pas **Hai(fisch) m**
 - ovčarski pas **Schäferhund m**

Navigation buttons at the bottom include 'Auswahl speichern' and 'Alle auswählen'.

The screenshot shows the Langenscheidt Online-Wörterbuch interface. The search bar contains 'igra' and the search button is labeled 'Suchen'. The results are categorized by source:

- 11 Treffer** (11 hits) from **Dictionary**:
 - igra** (gen pl igāra) **Spiel n**
 - igra** na srčcu oder hazardnā **igra Glücksspiel n**
 - igra riječi** **Wortspiel n**
 - igra skrivāča** **Versteckspiel n**
 - veselā **igra THEAT Lustspiel n**
- Dictionary**:
 - online-igra IT** **Onlinespiel n**
- Dictionary**:
 - rādio-igra** → **rādio-drama**

Knjižnica



KorAP

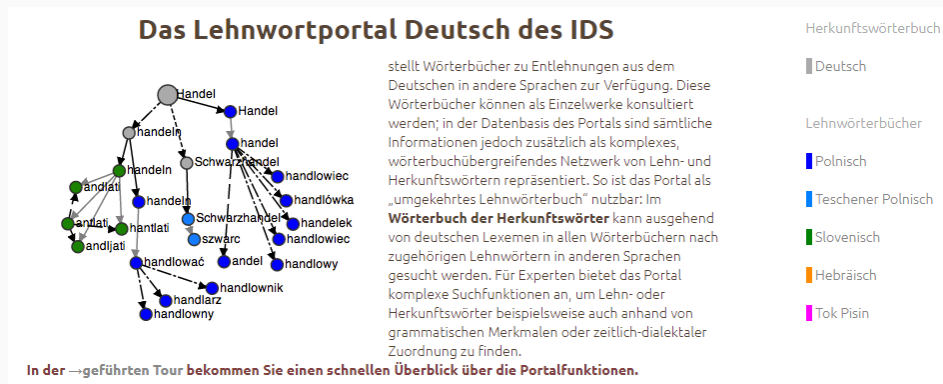
- KorAP (<https://korap.ids-mannheim.de/>) skalabilan je, fleksibilan i održiv sustav otvorenoga koda za rad s korpusima
- tražilica na kojoj se može pretraživati riječi te koristiti se upitima za pretraživanje
- trenutačno radi samo za njemački jezik, ali je u planu da se razvije i za pretraživanje drugih jezika





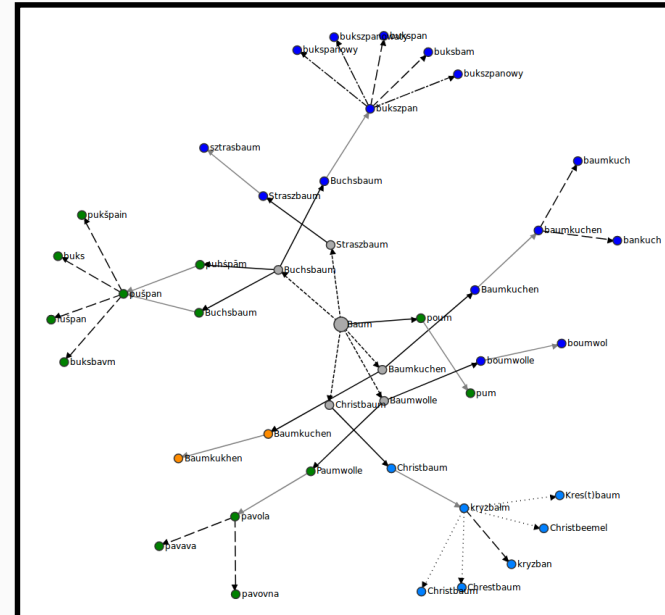
Das Lehnwortportal Deutsch des IDS

- rječnik posuđenica (<http://lwp.ids-mannheim.de/>) koji prikazuje koje su riječi iz njemačkoga ušle u poljski, hebrejski, Tok pisin i slovenski te koja je razlika u njegovu značenju u odnosu na njemački
- odnos između posuđenica prikazan je grafikonom



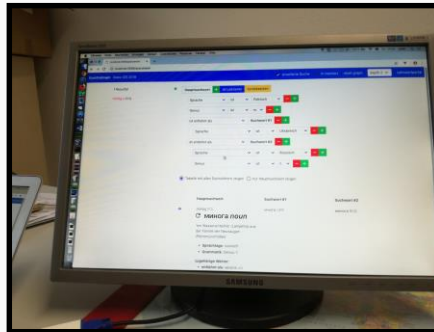
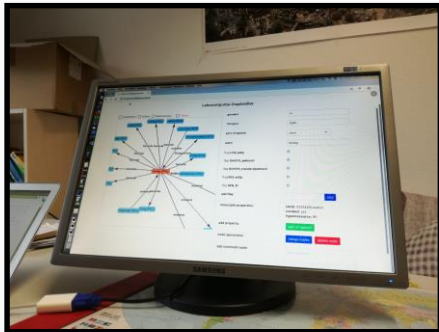
Das Lehnwortportal Deutsch des IDS

- odnos između posuđenica prikazan je grafikonom
- grafikoni su izrađeni s pomoću [D3 JavaScript](#) knjižnice
- podatci o riječima i vezama među riječima spremljeni su u bazi Oracle
- za unos i obradu natuknica koristi se XML
- stranica rječnika napravljena je s pomoću alata Play 2.7. i Spark



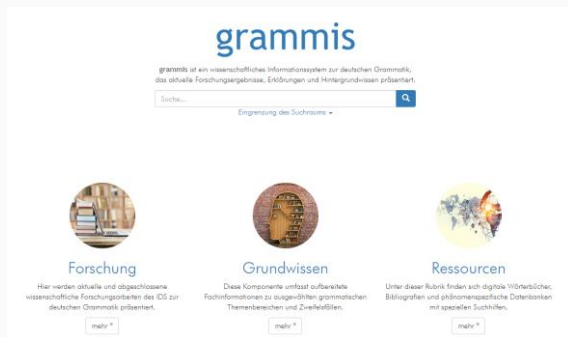
Das Lehnwortportal Deutsch des IDS

- trenutno se razvija novo sučelje koje će omogućiti paralelno pretraživanje više jezika odjednom
- u planu je da se dodaju posuđenice iz rumunjskoga i engleskoga
- novi grafikoni izrađuju se s pomoću jezika gremlin unutar alata [TinkerPop](#), u njima urednici mogu izravno uređivati metapodatke



grammis

- grammis je baza gramatičkih naziva (<https://grammis.ids-mannheim.de/>)
- na početnoj stranici osim tražilice nalaze se poveznice na resurse koji uključuju određene baze podataka (genitivi i pridjevi), popis priručnika (pravopisi i gramatike) i rječnike (govora, prijedloga, itd.)



grammis

- natuknice su obrađene u XML-u te pohranjene u bazi Oracle
- za izradu stranice koristio se CakePHP
- u budućnosti se planira napraviti virtualni pomagač na stranici koji će s pomoću glasovnih naredaba odvesti korisnika do željene poveznice na stranici

OWD

- OWD portal (<https://www.owid.de/>) prikuplja više rječnika i internetskih bibliografija (uglavnom za istraživačke svrhe) na jednome mjestu (digitalne i e-rječnike), koje je moguće pretraživati istodobno
- trenutačno razvijaju novo sučelje za portal jer staro nije responzivno
- noviji proizvodi koji se razvijaju na stranici mogu se vidjeti na poveznici: <https://www.owid.de/plus/index.html>

Wortspielereien

- virtualna stvarnost koji se pokreće s pomoću 3D naočala
- prikazan svemir u kojemu su riječi prikazane kao zvijezde
- kroz to zvijezde može se navigirati s pomoću posebnoga pokazivača
- postoji i igra unutar ove virtualne stvarnosti u kojoj igrači dobivaju određenu riječ koju moraju dovući do odgovarajuće skupine riječi na nebu
- program Unity s pomoću kojega se izrađuje
- demonstracija: <https://www.youtube.com/watch?v=YjaA2VGN81c>

Hvala na pozornosti!